Notations: Let $(X_i, \Delta_i), i = 1, \cdots, n$, be a sample of i.i.d. censored survival data, where $X_i = T_i \wedge C_i$ and $\Delta_i = I(T_i \leq C_i)$. We assume independence censoring assumption, i.e. $T_i \perp C_i$. In addition, we assume that all the failure times and censoring times are continuous. Define the counting process $N_i(t) = I(X_i \leq t, \Delta_i = 1)$ and and the at-risk process $Y_i(t) = I(X_i \geq t)$.

1. A very famous result given by Richard Gill allowed the Kaplan-Meier estimator to be expressed as a stochastic integral of a counting process martingale. This allows the application of martingale theory directly on the Kaplan-Meier estimator rather having to relate it to the Nelson Aalen estimator. To be specific,

$$\frac{\hat{S}_{KM}(t) - S(t)}{S(t)} = - \int_0^t \frac{\hat{S}_{KM}(u^-)}{S(u)} \frac{dM(u)}{Y(u)}, \tag{1}$$

where $S(t) = P(T \geq t)$ and is assumed to be continuous, $\hat{S}_{KM}(t)$ is right-continuous Kaplan-Meier estimator, $\hat{S}_{KM}(t^-)$ is left-continuous version of the Kaplan-Meier estimator, $Y(u) = \sum_{i=1}^n Y_i(u)$, $N(u) = \sum_{i=1}^n N_i(u)$, and $M(t) = N(t) - \int_0^t Y(u) \lambda(u) du$ with $\lambda(u)$ being the hazard function of failure time $T$.

a. Prove the equality (1).

b. Assume that $\sup_{t \in [0, \tau]} |\hat{S}_{KM}(t^-) - S(t)| \to 0$ in probability, as $n \to \infty$, where $P(X \geq \tau) \geq \epsilon > 0$. Based on equation (1), apply the martingale central limit theorem to derive the weak convergence of the stochastic process $\sqrt{n} \{\hat{S}_{KM}(t) - S(t)\}$ on $[0, \tau]$. (Note: You need to check that the conditions for the martingale CLT hold)

c. Find a consistent estimator for the asymptotic variance of $\sqrt{n} \{\hat{S}_{KM}(t) - S(t)\}$ and show that it is consistent.

2. Consider Problem 2 in Homework 2. Define $U_n(t) = n^{-3/2} Z^*(t)$.

a. Show that $U_n^*(t) \to N(0, \sigma^2(t))$ in distribution as $n \to \infty$ and find $\sigma^2(t)$. 

b. Find a consistent estimator for $\sigma^2(t)$ and show that it is consistent.